# Weekly Report

6/20/2016-6/26/2016

## Work

- I finish two exams this week.
- I complete two page poster for VIS2016.
- I read several papers about document clustering by means of matrix factorization.

## Plan for next week

- I need to prepare a course representation next week.

## Matrix Factorization

对于文档聚类[4]等问题，是把高维数据进行聚类的方法，也非常类似于降维，即一个文档可以有多个文档聚类（主题）来表达。给定数据矩阵$X = [x_{ij}] = [x_1, ..., x_n] \in \mathbb{R}^{m\times n}$，$X$的每列是样本向量。非负矩阵分解（NMF）[2]是一种致力于分析数据矩阵元素都非负的矩阵分解方法。NMF的目标是寻找两个非负矩阵$U = [u_{ij}] \in \mathbb{R}^{m\times t}$和$V = [v_{ij}] \in \mathbb{R}^{n\times t}$用来最小化下列目标函数：

$$O = \sum_{i,j}(x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij}) \tag{1}$$

$U$矩阵可以看做是文档的$t$个类，$V$矩阵表示的是文档以这$t$的类为基的线性组合。

$$x_j \approx \sum_{k=1}^{t} u_k v_{jk} \tag{2}$$

$$\tag{3}$$

其中$u_k$是$U$的第$k$列向量。因此每个数据向量$x_j$用$U$的列的线性组合表示，用$V$的成分进行加权。因此$U$可以认为包含了能够优化对X中数据进行线性近似的基。令$z_j = [v_{j1}, ..., v_{jk}]^t$，$z_j$可以是每个数据点用$U$中新的基进行的新表示。因为只用了很少的基来表示大量的数据向量，所以只有当基能够发现数据潜在结构的时候才会有好的近似[3]。

尽管等式1的目标函数关于$U$和$V$分别是凸的，但关于$UV$不是凸的。因此期望找到寻找全局最优的算法是不现实的。Lee Seung[3]提出了如下的迭代更新的算法：

$$\begin{aligned} u_{ik} &\leftarrow u_{ik} \frac{\sum_j (x_{ij} v_{jk} / \sum_k u_{ik} v_{jk})}{\sum_j v_{jk}} \\ v_{jk} &\leftarrow v_{jk} \frac{\sum_i (x_{ij} u_{ik} / \sum_k u_{ik} u_{jk})}{\sum_j u_{ik}} \end{aligned} \tag{4}$$

公式4可以证明上面的更新步骤会找到目标函数的局部最优解[3]。

NMF寻找基用来优化原始数据的线性近似。人们希望能够通过寻找更好的基发现数据的几何结构。一种自然的假设是如果两个数据$x_j$，$x_s$在数据分布的内在几何结构中足够近，那么在新的基下表示的$z_j$和$z_s$也足够近。 LPNMF[1]考虑有$N$个顶点的图，每个顶点对应语料库中的一个文本。如下定义权重矩阵：

$$W_{js} = \begin{cases} 1, & if \quad x_j \in N_p(x_s) \quad or \quad x_s \in N_p(x_j) \\ 0, & otherwise \end{cases} \tag{5}$$

其中，$N_p(x_s)$代表$x_s$的$p$最近邻的集合。

我们用两个数据在新基上的低维表示的散度来刻画“距离”：

$$D(z_j || z_s) = \sum_{k=1}^{t} (v_{jk} \log \frac{v_{jk}}{v_{sk}} - v_{jk} + v_{sk}) \tag{6}$$

因为我们有$z_j = [v_{j1}, ..., v_{jk}]^t$。因此，下面的能够用来度量低维表示沿数据内在几何结构侧底线变化的光滑性。

$$\begin{aligned} \mathcal{R} &= \frac{1}{2} \sum_{j,s=1}^{n} (D(z_j || z_s) + D(z_s || z_j)) W_{js} \\ &= \frac{1}{2} \sum_{j,s=1}^{n} \sum_{k=1}^{t} (v_{jk} \log \frac{v_{jk}}{v_{sk}} + v_{sk} \log \frac{v_{sk}}{v_{jk}}) W_{js} \end{aligned} \tag{7}$$

通过最小化R，我们获得了数据流形上足够光滑的条件概率分布。最小化R的一种直观解释是，如果$x_j$和$x_s$很近（也就是说$W_{js}$很大），$z_j$和$z_s$也离的很近。

给定数据矩阵$X = [x_{ij}] \in \mathbb{R}^{m \times n}$，LPNMF方法的目标是找到两个非负矩阵$U$和$V$，通过最小化了下面的目标函数：

$$O = \sum_{i,j} (x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij}) + \lambda \mathcal{R} \tag{8}$$

其中$Y = [y_{ij}] = UV^T$，$\lambda > 0$是正则参数。

LPNMF的计算复杂度是$O(n^3)$，其中n是样本数。加速局部保形非负矩阵分解[5]（Accelerated LPNMF， A-LPNMF）通过选择$p(p \ll n)$个锚点的方式将数据稀疏的表示成锚点的线性组合，解决了LPNMF的计算复杂性问题。 给定了$n$个样本$x_1, ..., x_n \in \mathbb{R}^m$，我们试图通过将数据$x_i$压缩为几个锚点的线性表示。换句话说，对任意数据点$x_i$，我们的目标是找到它的近似表示$\hat{x_i}$：

$$\hat{x_i} = \sum_{j=1}^{p} z_{ji} l_j \tag{9}$$

其中$\{l_j\}_{j=1}^p$是$p$个锚点。一族好的锚点能够很好的涵盖整个数据集。因此我们用Kmeans算法来生成锚点（把聚类中心当作锚点）。然后我们用$z_i = [z_{1i}, ..., z_{li}]^T$作为$x_i$的压缩。

一个自然的假设是如果$L_j$靠近$x_j$的话$z_{ji}$应该更大。如果$L_j$不在$x_i$的$r(<= l)$近邻内，我们可以通过将$z_{ji}$置为0来强化上述假设。这种限制自然的导致了稀疏压缩。 令$N_{(i)}$表示包含了$x_i$锚点的$r$最近邻的指标集，也就是说，如果$j \in N_{(i)}$，那么$l_j$在$x_i$锚点的$r$最近邻之间。我们依照下面来计算$z_{ji}$：

$$z_{ji} = \begin{cases} \frac{K(x_i, l_j)}{\sum_{j' \in N_{(i)}} K(x_i, l_{j'})}, & j \in N_{(i)} \\ 0, & j \notin N_{(i)} \end{cases} \tag{10}$$

其中$K(\cdot)$是核函数。我们可以简单的选取最常用的高斯核：

$$K(x_i, l_j) = exp(-\frac{||x_i - l_j||^2}{2\sigma^2}) \tag{11}$$

其中，sigma是核宽参数。

这一步计算在我们的AMTG转移矩阵的计算中也用到了，对轨迹片段聚类成$K$个类，然后每个轨迹片段可以表达为其相近的$c$个聚类的距离向量。看起来这样的方法在机器学习中经常用到，可以用于对高维数据的压缩，加快计算效率。这点在原始文献Storyline中应该是体现的比较明显，因为对图片抽取了几千维的特征，而我们AMTG原来的轨迹片段的特征维度也就10维。所以另一方面，我认为可能是为了是的计算转移比较有意义，是在不同聚类之间的状态转移，而不是在某一个特征上的转移（如，图像像素）。

# References

[1] Deng Cai, Xiaofei He, Xuanhui Wang, Hujun Bao, and Jiawei Han. Locality preserving nonnegative matrix factorization. In IJCAI, volume 9, pages 1010 – 1015, 2009.

[2] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755):788 – 791, 1999.

[3] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, pages 556 – 562, 2001.

[4] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pages 267 – 273. ACM, 2003.

[5] Guanhong Yao and Cai Deng. Accelerating locality preserving nonnegative matrix factorization. In Proceedings of the 21st ACM international conference on Information and knowledge management, pages 2271 – 2274. ACM, 2012.